

CLAIMS

1. A computerized method for predicting the correct addressee to be filled-in in an addressee field in an e-mail system, whereby user-related history information, including the user's sent and/or received e-mail, is analyzed for associating the most probable addressee for an e-mail to be addressed, comprising the steps of:

analyzing at least one of the following attributes of said user-related history information and of said e-mail to be addressed:

a subject line of said sent, received, and to-be-addressed e-mail;

the length of said sent, received, and to-be-addressed e-mail;

the language used in said sent, received, and to-be-addressed e-mail;

a time associated with said sent, received, and to-be-addressed e-mail;

the vocabulary used in said sent, received, and to-be-addressed e-mail;

topics discussed in the body of said sent, received, and to-be-addressed e-mail;

the salutation form used in said sent, received, and to-be-addressed e-mail;

the closing form used in said sent, received, and to-be-addressed e-mail;

whereby Text Mining methods are used to associate attribute values with respective addressees, thus yielding a plurality of single analysis results usable for said prediction, and

weighting the plurality of said single analysis results to provide a Data Mining Model adapted to offer at least one top favorite addressee proposal as a prediction result.

2. The method according to claim 1, wherein said e-mail system has multiple use modes, further comprising the step of:

using separate Data Mining models for different use modes.

3. The method according to claim 1, further comprising the step of:
performing a training of the Data Mining model triggered by any of the following criteria:
when a user overwrites the addressee proposal(s) made by the e-mail system, more frequently than limited by a predefined threshold level;
when the e-mail system is confronted with a number of new addressees not found in the user-related history information, and the number or fraction thereof is higher than a predefined threshold level;
after a predefined time limit has passed.
4. The method according to claim 1, in which the analysis results are generated in a table-like form, in which each attribute to be analyzed is associated with a predicted value, accompanied by a respective confidence value.
5. A computerized method for completing the addressee field in a user-initiated "new mail" within an e-mail system, comprising the steps of:
on an occurrence of an incomplete entering of an addressee term in said addressee field, running a predictive Data Mining method based on a trained Data Mining Model developed by:
analyzing one or more attribute values of user-related history information and of said e-mail to be addressed, whereby Text Mining methods are used to associate attribute values with respective addressees, thus yielding a plurality of single analysis results usable for said completing the address field; and
weighting the plurality of single analysis results to provide a Data Mining Model adapted to offer at least one top favorite addressee proposal as an address-

completion result; and

conveying at least the most probable addressee proposal to the user as an address-completion result.

6. The method of claim 5, wherein said analyzing step includes at least the step of analyzing one or more of the following attribute values of said user-related history:

- a subject line of sent, received, and/or to-be-addressed e-mail;
- the length of said sent, received, and/or to-be-addressed e-mail;
- the language used in said sent, received, and/or to-be-addressed e-mail;
- a time associated with said sent, received, and/or to-be-addressed e-mail;
- the vocabulary used in said sent, received, and/or to-be-addressed e-mail;
- topics discussed in the body of said sent, received, and/or to-be-addressed e-mail;
- the salutation form used in said sent, received, and/or to-be-addressed e-mail; and
- the closing form used in said sent, received, and/or to-be-addressed e-mail.

7. The method according to claim 5, further comprising the step of:
offering a subset of a predefined quantity of top favorite addressee proposals to the user for selection.

8. The method according to claim 5, further comprising the step of automatically providing an addressee field pre-filled with a top favorite addressee term.

9. The method according to claim 5, further comprising the step of
testing the Data Mining model on a test set of e-mail, said test set not being used in the attribute-value analyzing step, before predicting the most probable addressee, and

issuing a hint to the user, indicating the confidence of the predicted addressee proposal.

10. The method according to claim 5, further comprising the step of:
automatically expanding a trunk of an address term with the most probable addressee term, when the trunk of the most probable addressee term is present in the addressee field.

11. The method according to claim 5, further comprising the step of:
cross-checking an addressee term entered by the user with a list of top favorite addressees, determined by the system, and issuing a warning, if the probability is high that the user-entered addressee term is faulty.

12. A computerized system for predicting the correct addressee to be filled-in in an addressee field in an e-mail system, whereby user-related history information, including the user's sent and/or received e-mail, is analyzed for associating the most probable addressee for an e-mail to be addressed, comprising:

means for analyzing at least one of the following attributes of said user-related history information and of said e-mail to be addressed:

- a subject line of sent, received, and/or to-be-addressed e-mail;
- the length of said sent, received, and/or to-be-addressed e-mail;
- the language used in said sent, received, and/or to-be-addressed e-mail;
- a time associated with said sent, received, and/or to-be-addressed e-mail;
- the vocabulary used in said sent, received, and/or to-be-addressed e-mail;
- topics discussed in the body of said sent, received, and/or to-be-addressed e-mail;

the salutation form used in said sent, received, and/or to-be-addressed e-mail;
the closing form used in said sent, received, and/or to-be-addressed e-mail;
whereby Text Mining methods are used to associate attribute values with respective addressees, thus yielding a plurality of single analysis results usable for said prediction, and
means for weighting the plurality of said single analysis results to provide a Data Mining Model adapted to offer at least one top favorite addressee proposal as a prediction result.

13. The system according to claim 12, wherein said e-mail system has multiple use modes, further comprising:

means for using separate Data Mining models for different use modes.

14. The system according to claim 12, further comprising:

means for performing a retraining of the Data Mining model triggered by any of the following criteria:

when a user overwrites the addressee proposal(s) made by the e-mail system, more frequently than limited by a predefined threshold level;

when the e-mail system is confronted with a number of new addressees not found in the user-related history information, and the number or fraction thereof is higher than a predefined threshold level;

after a predefined time limit has passed.

15. The system according to claim 12, in which the analysis results are generated in a table-like form, in which each attribute to be analyzed is associated with a predicted value, accompanied by a respective confidence value.

16. A computerized system for completing the addressee field of a new e-mail being sent by a user via an e-mail system, comprising:

means for developing a trained Data Mining Model, comprising

means for analyzing one or more attribute values of user-related history information and of said new e-mail whereby Text Mining methods are used to associate attribute values with respective addressees, thus yielding a plurality of single analysis results usable for completing the address field of said new e-mail, and

means for weighting the plurality of single analysis results to provide a Data Mining Model adapted to identify at least the most probable addressee proposal from among the plurality of single analysis results as an address-completion result; and

means for conveying at least the most probable addressee proposal to the user as a proposed address-completion result;

whereby, on an occurrence of an incomplete entering of an addressee term in said addressee field, running said predictive Data Mining model to identify the most probable addressee proposal.

17. The system of claim 16, wherein said means for analyzing includes at least means for analyzing one or more of the following attribute values of said user-related history:

a subject line of sent, received, and/or said new e-mail;

the length of said sent, received, and/or new e-mail;
the language used in said sent, received, and/or new e-mail;
a time associated with said sent, received, and/or new e-mail;
the vocabulary used in said sent, received, and/or new e-mail;
topics discussed in the body of said sent, received, and/or new e-mail;
the salutation form used in said sent, received, and/or new e-mail; and
the closing form used in said sent, received, and/or new e-mail.

18. The system according to claim 16, further comprising:
means for presenting a subset of said single analysis results to said user, said subset comprising the most predefined quantity of top favorite addressee proposals to the user.
19. The system according to claim 16, further comprising means for automatically providing an addressee field pre-filled with the top favorite addressee.
20. The system according to claim 16, further comprising:
means for testing the Data Mining model on a test set of mails, not being part of the training step before predicting the most probable addressee, and
means for issuing a hint to the user, indicating the confidence of the predicted addressee proposal.
21. The system according to claim 16, further comprising:
in case of a trunk of the addressee term being present in the addressee field, and in case a high significance of the predictable addressee being present provided by the run of

the Data Mining method,

means for automatically expanding said trunk with the most probable addressee term.

22. The system according to claim 16, further comprising:

means for cross-checking a term entered by the user with a list of top favorite addressees, determined by the system, and issuing a warning, if the probability is high that the user-entered term is faulty.

23. A computer program product for predicting the correct addressee to be filled-in in an addressee field in an e-mail system, whereby user-related history information, including the user's sent and/or received e-mail, is analyzed for associating the most probable addressee for an e-mail to be addressed, the computer program product comprising a computer-readable storage medium having computer-readable program code embodied in the medium, the computer-readable program code comprising:

computer-readable program code that analyzes at least one of the following attributes of said user-related history information and of said e-mail to be addressed:

- a subject line of a sent, received, and/or to-be-addressed e-mail;
- the length of said sent, received, and/or to-be-addressed e-mail;
- the language used in said sent, received, and/or to-be-addressed e-mail;
- a time associated with said sent, received, and/or to-be-addressed e-mail;
- the vocabulary used in said sent, received, and/or to-be-addressed e-mail;
- topics discussed in the body of said sent, received, and/or to-be-addressed e-mail;
- mail;
- the salutation form used in said sent, received, and/or to-be-addressed e-mail;
- the closing form used in said sent, received, and/or to-be-addressed e-mail;

whereby Text Mining methods are used to associate attribute values with respective addressees, thus yielding a plurality of single analysis results usable for said prediction, and

computer-readable program code that weights the plurality of said single analysis results to provide a Data Mining Model adapted to offer at least one top favorite addressee proposal as a prediction result.

24. The computer program product according to claim 23, wherein said e-mail system has multiple use modes, further comprising:

computer-readable program code that uses separate Data Mining models for different use modes.

25. The computer program product according to claim 23, further comprising: computer-readable program code that performs a retraining of the Data Mining model triggered by any of the following criteria:

when a user overwrites the addressee proposal(s) made by the e-mail system, more frequently than limited by a predefined threshold level;

when the e-mail system is confronted with a number of new addressees not found in the user-related history information, and the number or fraction thereof is higher than a predefined threshold level;

after a predefined time limit has passed.

26. The computer program product according to claim 23, in which the analysis results are generated in a table-like form, in which each attribute to be analyzed is associated with a predicted value, accompanied by a respective confidence value.

27. A computer program product for completing the addressee field in a user-initiated e-mail to be addressed within an e-mail system, the computer program product comprising a computer-readable storage medium having computer-readable program code embodied in the medium, the computer-readable program code comprising:

computer-readable program code that, on an occurrence of an incomplete entering of an addressee term in said addressee field, runs a predictive Data Mining method based on a trained Data Mining Model developed by:

analyzing one or more attribute values of user-related history information and of said e-mail to be addressed,

whereby Text Mining methods are used to associate attribute values with respective addressees, thus yielding a plurality of single analysis results usable for said completing the address field, and

weighting the plurality of single analysis results to provide a Data Mining Model adapted to offer at least one top favorite addressee proposal as an address-completion result; and

computer-readable program code that conveys at least the most probable addressee proposal to the user as an address-completion result.

28. The computer program product of claim 27, wherein said computer-readable program code that analyzes includes at least computer-readable program code that analyzes one or more of the following attribute values of said user-related history:

a subject line of a sent, received, and/or to-be-addressed e-mail;

the length of said sent, received, and/or to-be-addressed e-mail;

the language used in said sent, received, and/or to-be-addressed e-mail;

a time associated with said sent, received, and/or to-be-addressed e-mail;
the vocabulary used in said sent, received, and/or to-be-addressed e-mail;
topics discussed in the body of said sent, received, and/or to-be-addressed e-mail;
the salutation form used in said sent, received, and/or to-be-addressed e-mail; and
the closing form used in said sent, received, and/or to-be-addressed e-mail.

29. The computer program product according to claim 27, further comprising:
computer-readable program code that offers a subset of predefined quantity of top
favorite addressee proposals to the user.

30. The computer program product according to claim 27, further comprising
computer-readable program code that automatically provides an addressee field pre-filled
with the top favorite addressee.

31. The computer program product according to claim 27, further comprising:
computer-readable program code that tests the Data Mining model on a test set of
mails, not being part of the training step before predicting the most probable addressee,
and
computer-readable program code that issues a hint to the user, indicating the
confidence of the predicted addressee proposal.

32. The computer program product according to claim 27, further comprising:
in case of a trunk of the addressee term being present in the addressee field, and in
case a high significance of the predictable addressee being present provided by the run of
the Data Mining method,

computer-readable program code that automatically expands said trunk with the most probable addressee term.

33. The computer program product according to claim 27, further comprising:
computer-readable program code that cross-checks a term entered by the user with a list of top favorite addressees, determined by the system, and issues a warning, if the probability is high that the user-entered term is faulty.